

# Where does Cosmological Perturbation Theory Break Down?

Cristian Armendariz-Picon, Michele Fontanini, Riccardo Penco, and Mark Trodden

*Department of Physics, Syracuse University, Syracuse, NY 13244, USA*

## Abstract

We apply the effective field theory approach to the coupled metric-inflaton system, in order to investigate the impact of higher dimension operators on the spectrum of scalar and tensor perturbations in the short-wavelength regime. In both cases, effective corrections at tree-level become important when the Hubble parameter is of the order of the Planck mass, or when the physical wave number of a cosmological perturbation mode approaches the square of the Planck mass divided by the Hubble constant. Thus, the cut-off length below which conventional cosmological perturbation theory does not apply is likely to be much smaller than the Planck length. This has implications for the observability of “trans-Planckian” effects in the spectrum of primordial perturbations.

## I. INTRODUCTION

One of the main successes of inflation [1] is the explanation of the origin of structure [2]. During slow-roll, the Hubble radius remains nearly constant, while cosmological modes are constantly pushed out of the horizon. Thus, local processes determine the amplitude and properties of perturbations at sub-horizon scales, which are transferred to cosmologically large distances by the accelerated expansion. In that sense, the sky is the screen upon which inflation has projected the physics of the microscopic universe.

The primordial perturbations seeded during inflation arise from quantum-mechanical fluctuations of the inflaton around its homogeneous value. Hence, their properties directly depend on the quantum state of the inflaton perturbations. Conventionally, this is taken to be a state devoid of quanta in the asymptotic past, raising the crucial question of whether we can trust cosmological perturbation theory—and its quantum nature—at such early times [3].

According to our present understanding, quantum field theories and general relativity are merely low energy descriptions of a more fundamental theory of quantum gravity. In the case of inflation, the leading terms in the corresponding effective Lagrangian are the Einstein-Hilbert term plus the inflaton kinetic term and potential. In an effective field theory treatment, these terms are accompanied by all other possible operators compatible with the symmetries of the theory, namely, general covariance and any other symmetry of the inflaton sector. Higher dimensional operators are suppressed by powers of an energy scale, which we will assume to be of the order of the reduced Planck mass  $M_p \equiv (8\pi G)^{-1/2}$ , and they are therefore expected to be negligible at sufficiently small momenta, or sufficiently long wavelengths. The goal of this paper is to determine the three-momentum scale  $\Lambda$  at which these operators become important. The terms that yield the leading (momentum-independent) corrections to the primordial spectrum have been recently discussed in [4].

It is crucial to realize that  $\Lambda$  does not need to equal the Planck scale. As an extreme example, let us consider perturbations around a background given by a constant scalar field in Minkowski spacetime,  $\varphi = \varphi_0 + \delta\varphi$  and  $g_{\mu\nu} = \eta_{\mu\nu} + \delta g_{\mu\nu}$ . In this case, the effective action for the perturbations is Lorentz invariant, because the background is Lorentz invariant and we assume that the original effective action  $S[\varphi, g_{\mu\nu}]$  is generally covariant (we are implicitly assuming that a Lorentz-invariant gauge on  $\delta g_{\mu\nu}$  has been imposed.) Therefore, the regime in which higher order operators become important will be expressed by Lorentz invariant

relations such as, for example,  $k_\mu k^\mu \approx M_p^2$ . In other words, there is not any bound on the three-momentum  $|\mathbf{k}|$ , simply because a relation like  $|\mathbf{k}| \approx M_p$  is not Lorentz invariant. One would like to extend this argument to an expanding universe. On short time-scales and short distances, spacetime can be regarded as flat and the scalar field driving inflation can be regarded as constant. Hence, it seems that cosmological perturbation theory should be valid all the way into the regime  $|\mathbf{k}| \rightarrow \infty$ , because the relevant four-momenta of cosmological perturbations are light-like,  $k_\mu k^\mu = 0$ . As we shall see though, the evolution of the inflaton leads to small but finite violations of the Lorentz symmetry even in the short-wavelength limit, which are imprinted on the power spectrum as  $\mathbf{k}$ -dependent corrections.

The phenomenological imprints of trans-Planckian physics on the primordial spectrum of perturbations, and the implications of a finite cut-off  $\Lambda$  on the spatial momentum of cosmological modes have been extensively studied [5–30]. These articles mostly study corrections to the power spectrum in the long-wavelength limit  $k/a \equiv k_{ph} \ll H$ , at late times, which is the regime directly accessible by experimental probes. In this article we focus instead on the short-wavelength regime  $k_{ph} \gg H$ , at early times, since we are interested in determining how far into the ultraviolet cosmological perturbation theory applies. At short wavelengths, the power spectrum can be cast as a derivative expansion of the form

$$\langle \delta\varphi^*(\mathbf{k})\delta\varphi(\mathbf{k}) \rangle = \frac{1}{2k} \left( 1 + \alpha_2 \frac{k_{ph}^2}{M_p^2} + \alpha_4 \frac{k_{ph}^4}{M_p^4} + \dots \right), \quad (1.1)$$

with coefficients  $\alpha_i$  that depend on slow-roll parameters and the dimensionless ratio  $H/M_p$ . The analytic corrections to the leading result  $1/2k$  arise from tree-level diagrams with vertices from higher-dimensional operators. We only consider tree-level diagrams here, since we expect loop diagrams to simply introduce a logarithmic dependence of the dimensionless coefficients  $\alpha_i$  on scale. Cosmological perturbation theory fails when the expansion in powers of  $k_{ph}$  breaks down, namely, when all the terms become of the same order,

$$k_{ph} \approx M_p \sqrt{\frac{\alpha_{2n}}{\alpha_{2n+2}}} \equiv \Lambda. \quad (1.2)$$

As we shall show, the ratios  $\alpha_{2n}/\alpha_{2n+2}$  are all quite large and of the same order, so the effective cut-off  $\Lambda$  significantly differs from  $M_p$ . In a slightly different context, a similar analysis has been applied to the bispectrum in [31].

The structure of this paper is as follows. In the next section we describe the relevant background to our problem, setting up a description of perturbation theory in cosmology and

describing the basics of tensor and scalar fluctuations. In section III we compute the squared amplitude of tensor perturbations and we derive the results mentioned above. In section IV we apply a similar analysis to the case of scalar perturbations, and obtain similar results. We conclude and discuss possible implications of our results in section V. Throughout this paper we work in units such that  $\hbar=c=1$ , and with metric signature  $(- +++)$ .

## II. SETTING THE SCENE – COSMOLOGICAL PERTURBATION THEORY

### A. The Inflating Background

Our starting point is a standard single-field inflation model. At sufficiently late times, the inflaton and gravity must be described by a low-energy effective action, whose leading terms are dictated by general covariance and the field content,

$$S_0 = \int d^4x \sqrt{-g} \left[ \frac{M_p^2}{2} R - \frac{1}{2} \partial_\mu \varphi \partial^\mu \varphi - V(\varphi) \right]. \quad (2.1)$$

In an effective field theory context, the action should also contain additional terms suppressed by powers of a dimensionful scale, which we assume to be of the order of the Planck mass  $M_p$ . Our goal is to determine the point beyond which such higher-dimensional operators produce corrections to the two-point function of cosmological perturbations that cannot be neglected. Our considerations can be readily generalized to cases in which the suppression scale of the higher-dimensional operators is not the Planck mass, but any other scale.

If the potential  $V(\varphi)$  is sufficiently flat, at least in a certain region in field space, there exist inflationary solutions, along which a homogeneous scalar field  $\varphi(\eta, \mathbf{x}) = \varphi_0(\eta)$  slowly rolls down the potential and spacetime is spatially homogeneous, isotropic and flat,

$$g_{\mu\nu}^{(0)} \equiv a^2(\eta) \eta_{\mu\nu}, \quad (2.2)$$

where  $\eta_{\mu\nu}$  is the Minkowski metric and  $\eta$  denotes conformal time. A model-independent measure of the slowness of the inflation is given by the slow-roll parameter

$$\epsilon \equiv -\frac{H'}{aH^2}, \quad (2.3)$$

where  $H \equiv a'/a^2$  is the Hubble parameter and a prime denotes a derivative with respect to conformal time. During slow-roll,  $\epsilon$  is nearly constant, and to lowest order in slow-roll

parameters its time derivative can be neglected. Throughout this article we work to leading non-vanishing order in the slow-roll expansion.

## B. Cosmological Perturbations

Let us now consider cosmological perturbations around the homogeneous and isotropic background described above. Writing  $\varphi = \varphi_0 + \delta\varphi$  and  $g_{\mu\nu} = g_{\mu\nu}^{(0)}(\eta) + \delta g_{\mu\nu}(\eta, \mathbf{x})$ , and substituting into equation (2.1), we can expand the action  $S_0$  up to the desired order in the fluctuations  $\delta\varphi$  and  $\delta g_{\mu\nu}$ ,

$$S_0[\varphi, g_{\mu\nu}] = \delta_0 S_0 + \delta_1 S_0 + \delta_2 S_0 + \dots \quad (2.4)$$

The lowest order term  $\delta_0 S_0$  does not contain any fluctuations and describes the inflating background; the linear term  $\delta_1 S_0$  vanishes because it corresponds to the first variation of the action along the background solution, and the quadratic part of the action  $\delta_2 S_0$  describes the free dynamics of the perturbations. The latter is what we need in order to calculate the primordial spectrum of fluctuations. To quadratic order, tensor and scalar perturbations are decoupled, so we may study them separately.

### 1. Tensor Perturbations

Tensor perturbations are described by a transverse and traceless tensor  $h_{ij}$ ,

$$ds^2 = a^2(\eta) [-d\eta^2 + (\delta_{ij} + h_{ij})dx^i dx^j] \quad (2.5)$$

The tensor  $h_{ij}$  itself can be decomposed in plane waves of two different polarizations, which again decouple at quadratic order. We shall hence focus on just one of them,

$$h_{ij}(\eta, \mathbf{x}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} e_{ij}(\mathbf{k}) h_k(\eta) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (2.6)$$

where the  $h_k(\eta)$  are the corresponding mode functions, and  $e_{ij}(\mathbf{k})$  denotes the normalized graviton polarization tensor,  $e^i{}_j e^j{}_i = 1$  (we raise and lower spatial indices with the Kronecker delta.) Note that for later convenience we work in a toroidal universe of volume  $V = L^3$ ; hence, the spatial wave numbers have components  $k_i = n_i \cdot (2\pi/L)$ , where the  $n_i$  are arbitrary integers.

Substituting the expansion (2.6) into the action (2.1), and using the background equations of motion, we may then express the free action  $\delta_2 S_0$  as

$$\delta_2 S_0 = \frac{1}{2} \int d\eta \sum_{\mathbf{k}} \left[ v'_k v'_{-k} - \left( k^2 - \frac{a''}{a} \right) v_k v_{-k} \right] , \quad (2.7)$$

where the scalar variable  $v_k$  is defined as

$$v_k = a M_p h_k . \quad (2.8)$$

Thus, in terms of  $v_k$  the action for tensor perturbations takes the form of an harmonic oscillator with time-dependent frequency.

## 2. Scalar Perturbations

In spatially flat gauge, the perturbed metric reads

$$ds^2 = a^2(\eta) [-(1 + 2\phi)d\eta^2 + 2\partial_i B dx^i d\eta + \delta_{ij} dx^i dx^j] . \quad (2.9)$$

On first inspection there appear to be three independent scalar variables:  $\phi$ ,  $B$  and the inflaton perturbation  $\delta\varphi$ . However, Einstein equations impose constraints on both  $\phi$  and  $B$ . Solving the corresponding Fourier transformed equations to leading order in the slow-roll expansion, one finds (see e.g. [32])

$$\phi_k = \sqrt{\frac{\epsilon}{2}} \frac{\delta\varphi_k}{M_p} , \quad (2.10a)$$

$$B_k = \sqrt{\frac{\epsilon}{2}} \frac{\delta\varphi'_k}{M_p k^2} . \quad (2.10b)$$

Consequently, there is only one physical scalar degree of freedom, and scalar perturbations can be described by just one variable. A convenient choice that is particularly useful for quantizing scalar perturbations is the Mukhanov variable [33], which in spatially flat gauge takes the simple form

$$v_k = a \delta\varphi_k . \quad (2.11)$$

Using relations (2.10) and (2.11), we may express  $\delta_2 S_0$  in terms of  $v_k$  only. For constant  $\epsilon$ , that is, to leading order in the slow-roll expansion, the resulting action is also given by equation (2.7). This agreement greatly simplifies the analysis, because it allows us to use the same set of propagators to describe both scalar and tensor fluctuations.

To leading order in the slow-roll expansion, the mode functions of both scalar and tensor perturbations hence satisfy the same equation of motion during inflation. Varying the action (2.7) with respect to  $v_{-k}$  we obtain

$$v_k'' + \left[ k^2 - \frac{a''}{a} \right] v_k = 0 , \quad (2.12)$$

which has a unique solution for appropriate initial conditions. The conventional choice is the Bunch-Davies or adiabatic vacuum, whose mode functions obey

$$v_k(\eta) \xrightarrow{|k\eta| \gg 1} \frac{e^{-ik\eta}}{\sqrt{2k}} \left[ 1 + \mathcal{O}\left(\frac{1}{k\eta}\right) \right] . \quad (2.13)$$

Because we are only interested in the sub-horizon limit, this is all we need to know about the mode functions. In particular, because the behavior of the mode functions in the short-wavelength limit does not depend on the details of inflation, our results are also insensitive to the particular form of the inflaton potential.

### C. Quantum Fluctuations and the *in-in* Formalism

In order to study the properties of cosmological modes in the short-wavelength regime, we concentrate on the two-point function of the field  $v$ ,

$$\langle v^*(\eta, \mathbf{k}) v(\eta, \mathbf{k}) \rangle \equiv \langle 0, in | v^*(\eta, \mathbf{k}) v(\eta, \mathbf{k}) | 0, in \rangle , \quad (2.14)$$

where  $|0, in\rangle$  is the quantum state of the perturbations, which we assume to be the Bunch-Davies vacuum. The two-point function characterizes the mean square amplitude of cosmological perturbation modes, and differs from the power spectrum just by a normalization factor. Note that in an infinite universe, the two-point function is proportional to a momentum-conserving delta function, which in a spatially compact universe is replaced by a Kronecker delta.

In the *in-in* formalism (see [34] for a clear and detailed exposition) the two-point function can be expressed as a path integral,

$$\begin{aligned} \langle v^*(\eta, \mathbf{k}) v(\eta, \mathbf{k}) \rangle &= \\ &= \int \mathcal{D}v_+ \mathcal{D}v_- v_+^*(\eta, \mathbf{k}) v_-(\eta, \mathbf{k}) \exp(iS_{\text{free}}[v_+, v_-]) \exp(iS_{\text{int}}[v_+]) \exp(-iS_{\text{int}}[v_-]) , \end{aligned} \quad (2.15)$$

where  $S_{\text{free}}$  is quadratic in the fields, and  $S_{\text{int}}$  contains not just the remaining cubic and higher order terms in the action, but also any other quadratic terms we may decide to regard as

perturbations. Note that there are two copies of the integration fields  $v_-$  and  $v_+$ , because we are calculating expectation values, rather than *in-out* matrix elements. This path integral expression is very useful to perturbatively expand the expectation value in powers of any interaction. In particular, each contribution can be represented by a Feynman diagram, with vertices drawn from the terms in  $S_{\text{int}}$  and propagators determined by the free action  $S_{\text{free}}$ . In our case, the latter are given by

$$\bullet \text{---} \bullet = \int \mathcal{D}v_+ \mathcal{D}v_- v_+^*(\eta, \mathbf{k}) v_+(\eta', \mathbf{k}) \exp(iS_{\text{free}}) \approx \frac{e^{-ik|\eta-\eta'|}}{2k}, \quad (2.16a)$$

$$\bullet \text{---} \bullet = \int \mathcal{D}v_+ \mathcal{D}v_- v_-^*(\eta, \mathbf{k}) v_-(\eta', \mathbf{k}) \exp(iS_{\text{free}}) \approx \frac{e^{ik|\eta-\eta'|}}{2k}, \quad (2.16b)$$

$$\bullet \text{---} \bullet = \int \mathcal{D}v_+ \mathcal{D}v_- v_+^*(\eta, \mathbf{k}) v_-(\eta', \mathbf{k}) \exp(iS_{\text{free}}) \approx \frac{e^{ik(\eta-\eta')}}{2k}, \quad (2.16c)$$

which we quote here just in the sub-horizon limit. Note that to first order in  $S_{\text{int}}$  there are two vertices, one that contains powers of  $v_+$  and one that contains powers of  $v_-$ ; the associated coefficients just differ by an overall sign.<sup>1</sup>

As a simple example, let us calculate the value of the two-point function in the short-wavelength limit to zeroth order in the interactions. Using the definition (2.15) and equation (2.16c), we find

$$\langle v^*(\eta, \mathbf{k}) v(\eta, \mathbf{k}) \rangle = \bullet \text{---} \bullet \approx \frac{1}{2k} \quad (|k\eta| \gg 1), \quad (2.17)$$

which is the well-known and standard short-wavelength limit result. In this regime, the two-point function is hence the inverse of the dispersion relation, since the latter determines the appropriate boundary conditions for the mode functions, as in equation (2.13).

In the next two sections we use the path integral (2.15) to calculate the corrections to the two-point function coming from higher-order operators in the action. These can be interpreted as corrections to the dispersion relation, even though in the presence of such terms the mode equations generally contain higher order time derivatives. In any case, a significant disagreement between the calculated two-point function and the lowest order

<sup>1</sup> The quadratic action  $S_{\text{free}}$  enforces  $v_+(\vec{k}) = v_-(\vec{k})$  at time  $\eta$ . Hence, we could replace  $v_+^*(\eta, \mathbf{k}) v_-(\eta, \mathbf{k})$  by  $v_+^*(\eta, \mathbf{k}) v_+(\eta, \mathbf{k})$  or  $v_-^*(\eta, \mathbf{k}) v_-(\eta, \mathbf{k})$  inside the path integral (2.15). Our choice removes the apparently ill-defined corrections we otherwise obtain when higher-order time derivatives act on the time-ordered products in equations (2.16a) and (2.16b). These ill-defined corrections can also be eliminated by field-redefinitions, a procedure that leads to the same corrections we find using our choice of field insertions.

result (2.17) points to the lack of self-consistency of our quantization procedure, and signals the breakdown of cosmological perturbation theory.

### III. THE LIMITS OF PERTURBATION THEORY: TENSORS

The lowest order action (2.1) contains the leading terms that describe the dynamics of the inflaton and its perturbations. However, as we have noted, in an effective field theory approach the action generically contains all possible terms compatible with general covariance and any other symmetry of the theory. Here, for simplicity, we assume invariance under parity, an approximate shift symmetry of the inflaton, and a discrete  $\mathbb{Z}_2$  symmetry  $\varphi \rightarrow -\varphi$ . Thus, all possible effective corrections to the action (2.1) can be built from the metric  $g_{\mu\nu}$ , the Riemann tensor  $R_{\mu\nu\lambda\rho}$ , the covariant derivative  $\nabla_\mu$  and an even number of scalar fields  $\varphi$ . In what follows, we consider these additional terms and compute the corrections they induce on the two-point function of tensor perturbations in the short-wavelength limit. This allows us to determine the regime in which additional terms in the action cannot be neglected, and hence, the range over which cosmological perturbation theory is applicable. The reader not interested in technical details may skip directly to section III C, where we collect and summarize our results.

#### A. Dimension Four Operators

We begin our analysis by considering all dimension four operators, which appear in the action multiplied by dimensionless coefficients. On dimensional grounds, we expect these to yield corrections to the two-point function that are suppressed by only two powers<sup>2</sup> of  $M_p$ . These operators will also help us to illustrate our formalism and discuss some of the important issues related to our calculation.

Any generally covariant dimension four effective correction must be of the form

$$S_1 \equiv S_\alpha + S_\beta = \int \sqrt{-g} (\alpha R^2 + \beta C^2), \quad (3.1)$$

where  $C^2$  is the square of the Weyl tensor,

$$C^2 = R_{\mu\nu\lambda\rho} R^{\mu\nu\lambda\rho} - 2R_{\mu\nu} R^{\mu\nu} + \frac{1}{3} R^2, \quad (3.2)$$

---

<sup>2</sup> Dimension six operators quadratic in  $\varphi$  also contribute at this order; we consider them later.

and the dimensionless couplings  $\alpha$  and  $\beta$  are assumed to be of order one. Note that we have ignored total derivatives like the Gauss-Bonnet term, since they do not lead to any corrections in perturbation theory. The Levi-Civita tensor cannot appear in the action because we assume invariance under parity.

We start by substituting the perturbed metric (2.5) into equation (3.1) and expanding up to second order in  $h_{ij}$ . Using the modified background equations and equation (2.8) to express the tensor perturbations in terms of the variable  $v$ , we obtain in the sub-horizon limit

$$\begin{aligned}\delta_2 S_\alpha &= \frac{\alpha}{2M_p^2} \sum_{\mathbf{k}} \int d\eta' \left\{ -\frac{6a''}{a^3} v_k [v''_{-k} + k^2 v_{-k}] - \frac{6a''}{a^3} [v''_k + k^2 v_k] v_{-k} \right\}, \\ \delta_2 S_\beta &= \frac{\beta}{M_p^2} \sum_{\mathbf{k}} \int d\eta' \left\{ \frac{1}{a} [v''_k + k^2 v_k] - 2aH \left( \frac{v_k}{a} \right)' \right\} \cdot \left\{ \frac{1}{a} [v''_{-k} + k^2 v_{-k}] - 2aH \left( \frac{v_{-k}}{a} \right)' \right\}.\end{aligned}\quad (3.3)$$

From these expressions, it is easy to derive the rules for the vertices

$$\begin{aligned}\overline{\text{---}}^\alpha \times \overline{\text{---}} &\approx \frac{i\alpha}{M_p^2} \int_{-\infty}^\eta d\eta' \left\{ -\frac{6a''}{a^3} \left( \overrightarrow{\partial}_{\eta'}^2 + k^2 \right) - \left( \overleftarrow{\partial}_{\eta'}^2 + k^2 \right) \frac{6a''}{a^3} \right\} \\ \overline{\text{---}}^\alpha \times \overline{\text{---}} &= - \overline{\text{---}}^\alpha \times \overline{\text{---}}\end{aligned}\quad (3.4a)$$

$$\begin{aligned}\overline{\text{---}}^\beta \times \overline{\text{---}} &\approx \frac{2i\beta}{M_p^2} \int_{-\infty}^\eta d\eta' \left\{ \left( \overleftarrow{\partial}_{\eta'}^2 + k^2 \right) - \overleftarrow{\partial}_{\eta'} 2aH \right\} \frac{1}{a^2} \left\{ \left( \overrightarrow{\partial}_{\eta'}^2 + k^2 \right) - 2aH \overrightarrow{\partial}_{\eta'} \right\} \\ \overline{\text{---}}^\beta \times \overline{\text{---}} &= - \overline{\text{---}}^\beta \times \overline{\text{---}},\end{aligned}\quad (3.4b)$$

where the arrows indicate the propagator on which the derivatives act (because the vertex is quadratic, two propagators meet at the vertex.)

We are now ready to consider the correction due to the square of the Ricci scalar. The first order correction to the two-point function is given by the sum of the following two graphs,

$$\begin{aligned}\bullet \overline{\text{---}}^\alpha \times \overline{\text{---}}^\alpha \bullet &\approx \frac{i\alpha}{M_p^2} \int_{-\infty}^\eta d\eta' \left\{ \frac{i\delta(\eta - \eta')}{2k} \frac{6a''}{a^3} \right\} \approx -\frac{\alpha}{2k} \frac{12H^2}{M_p^2}, \\ \bullet \overline{\text{---}}^\alpha \times \overline{\text{---}}^\alpha \bullet &= \left( \bullet \overline{\text{---}}^\alpha \times \overline{\text{---}}^\alpha \bullet \right)^*,\end{aligned}\quad (3.5)$$

where we have used the fact that  $a''/a^3 \approx 2H^2$  to lowest order in slow-roll. Notice that the operator  $(\overrightarrow{\partial}_{\eta'}^2 + k^2)$  acting on the time-ordered propagators (2.16a) or (2.16b) produces a delta function, since both are Green's functions. On the other hand, when the same operator

acts on the propagator (2.16c) we get zero, because the latter is a regular solution of the free equation of motion (2.12) in the sub-horizon limit. This remark will turn out to be very useful when studying higher dimension operators.

We can now consider the correction due to the square of the Weyl tensor. In this case, the first order contribution is given by the sum of the following two graphs

$$\begin{aligned} \text{---} \bullet &= -\frac{2i\beta}{M_p^2} \int_{-\infty}^{\eta} d\eta' \left\{ \delta(\eta - \eta') aH + H^2 e^{2ik(\eta' - \eta)} \right\} \approx -\frac{\beta}{2k} \left\{ \frac{4iHk_{ph}}{M_p^2} + \frac{2H^2}{M_p^2} \right\} \\ \text{---} \bullet &= \left( \text{---} \bullet \right)^*. \end{aligned} \quad (3.6)$$

Note that the imaginary parts cancel once we sum the two graphs. This result is quite general and ensures that only corrections with even powers of  $k_{ph}$  appear.

In conclusion, we have found that the leading corrections due to dimension four operators result in a two-point function which in the short-wavelength limit has the form

$$\text{---} \bullet + \left( \text{---} \bullet + \text{---} \bullet + \text{c.c.} \right) \approx \frac{1}{2k} \left[ 1 - (24\alpha + 4\beta) \frac{H^2}{M_p^2} \right]. \quad (3.7)$$

Thus, when  $H$  becomes of order  $M_p$ , these corrections become as important as the leading result, and standard cosmological perturbation theory ceases to be applicable, as the reader may have expected.

## B. Higher Dimension Operators

We would now like to consider a generic operator of dimension  $2d + 4$ , suppressed by a factor of order  $1/M_p^{2d}$ . However, it turns out that considering directly corrections to the action (2.7) for the perturbations is a much more efficient approach than starting from generally covariant effective terms added to the Lagrangian (2.1), particularly if we are interested in identifying the dominant corrections in the sub-horizon limit. Hence, we shall focus directly on modifications to the action for the perturbations. A related approach has been described in [35].

Dimensional analysis implies that any operator of dimension  $2d + 4$ , quadratic in the dimensionless tensor perturbations  $h_{ij}$  and proportional to  $2f$  powers of the inflaton field  $\varphi$  must contain  $2d - 2f + 4$  derivatives  $\partial_\mu$  acting on  $h_{ij}$ ,  $\varphi_0(\eta)$  and  $a(\eta)$ . The derivatives can

be distributed and contracted using the Minkowski metric in many different ways,<sup>3</sup> but each of these terms can be schematically represented as

$$M_p^{-2d-2} (\partial^{2n+m+p} [a, \varphi_0]) (\partial^{2q+m+r} v) (\partial^{2s+p+r} v), \quad (3.8)$$

where we have used equation (2.8) and  $\partial^i[a, \varphi_0]$  is just a symbol that represents any combination of  $i$  derivatives acting on  $a$ 's and  $\varphi_0$ 's. One such term would have  $2n + m + p$  derivatives acting on one or more factors of  $a$  or  $\varphi_0$ ,  $2q + m + r$  derivatives acting on one field  $v$  and  $2s + p + r$  acting on the other  $v$ . In particular,  $2n$  of the derivatives acting on the scale factor or the background field are contracted among themselves while  $m$  and  $p$  of them are contracted with derivatives acting on, respectively, the first and second field  $v$ . The derivatives acting on the fields  $v$  are organized in a similar way.

Let us illustrate this notation by considering a term with  $p = s = f = 0$ ,  $m = n = q = 1$  and  $r = 2$ . Dimensional analysis implies that  $d = 3$ , and thus the explicit form of such a term would be

$$M_p^{-8} \partial_{2+1+0} [a] \partial_{2+1+2} v \partial_{0+0+2} v = M_p^{-8} \partial_\mu \partial^\mu \partial_\nu [a] \partial_\lambda \partial^\lambda \partial^\nu \partial_\alpha \partial_\beta v \partial^\alpha \partial^\beta v, \quad (3.9)$$

where  $\partial_\mu \partial^\mu \partial_\nu [a]$  denotes all possible ways to construct a term with three derivatives of the scale factor, with the given tensor structure.

The first step to estimate the leading correction due to a term of the form (3.8) is realizing that this can always be expressed as a linear combination of terms of the form

$$M_p^{-2d-2} \partial^{2j+l} [a, \varphi_0] \partial^{2m+l} v \partial_\mu \partial^\mu v, \quad (3.10)$$

plus, possibly, a term with no derivatives acting on  $v$ , which in any case gives a contribution that is always subdominant in the sub-horizon limit. For a proof that this decomposition is always possible, we refer the reader to Appendix A. In what follows, we therefore restrict ourselves to terms of the form (3.10).

<sup>3</sup> The reader may think that derivatives could be contracted not only among each other with the Minkowski metric, but also by using the additional tensor structure provided by the metric perturbations  $\delta g_{\mu\nu}/a^2 = h_{ij}\eta_{\mu i}\eta_{\nu j}$ . However, it turns out that  $(\delta g_{\mu\nu}/a^2) \partial^\nu a = h_{ij}\eta_{\mu i}\partial^j a = 0$  and, since  $h_{ij}$  is transverse,

$$\partial^\mu (\delta g_{\mu\nu}/a^2) = \eta_{\nu j}\partial^j h_{ij} = 0.$$

Thus, we get a non-vanishing contribution only when we contract derivatives with the Minkowski metric while the factors  $\eta_{\mu i}\eta_{\nu j}$  are contracted among each other yielding an irrelevant overall factor.

Dimensional analysis requires that the indexes  $j$ ,  $l$  and  $m$  in equation (3.10) obey

$$j + l + m = d - f + 1 . \quad (3.11)$$

Furthermore, since  $d^n\varphi_0/d\eta^n \propto \sqrt{2\epsilon} M_p a^n H^n$  and  $d^n a/d\eta^n \propto a^{n+1} H^n$ , each field  $\varphi_0$  yields a factor of  $M_p$ , while each derivative acting on it or on the scale factor results in a factor of  $H$  to leading order in slow-roll. Finally, the  $l$  partial derivatives  $\partial_\mu$  acting on  $v$  that are contracted with derivatives acting on  $a$  or  $\varphi_0$  can be turned into derivatives with respect to  $\eta$  only. Thus, (3.10) can be re-written as

$$M_p^{-2d+2f-2} f(a) H^{2j+l} \square^m \partial_\eta^l v \square v , \quad (3.12)$$

where we have defined  $\square \equiv \partial_\mu \partial^\mu$ ; the corresponding correction to the two-point function is schematically given by

$$\begin{aligned} \bullet \text{---} \times \text{---} \bullet &= \frac{i}{M_p^{2d-2f+2}} \int_{-\infty}^{\eta} d\eta' f(a) H^{2j+l} \times \\ &\times \bullet \text{---} \bullet \quad \left( \overleftarrow{\square}^m \overleftarrow{\partial}_{\eta'}^l \overrightarrow{\square} + \overrightarrow{\square} \overrightarrow{\partial}_{\eta'}^l \overrightarrow{\square}^m \right) \bullet \text{---} \text{---} \bullet \end{aligned} \quad (3.13)$$

plus the complex conjugate of this graph. Because (2.16c) satisfies the free equation of motion, this correction is non-vanishing only when the index  $m$  is equal to zero, and in this case we obtain

$$\bullet \text{---} \times \text{---} \bullet = \frac{1}{M_p^{2d-2f+2}} \int_{-\infty}^{\eta} d\eta' f(a) H^{2j+l} \delta(\eta - \eta') \frac{(ik)^l}{2k} = \frac{f(a)}{2k} \frac{H^{2j+l} (ik)^l}{M_p^{2d-2f+2}} . \quad (3.14)$$

The leading correction in the short-wavelength limit is the one with the maximum number of powers of  $k$ . According to equation (3.11), this maximum number simply equals  $d - f + 1 \equiv l_{max}$ , and it corresponds to the case in which  $j = m = 0$ . Thus, if  $d - f$  is odd,  $l_{max}$  is even and the leading correction is simply given by

$$\delta \langle v^*(\mathbf{k}) v(\mathbf{k}) \rangle \propto \frac{1}{2k} \left( \frac{H}{M_p} \right)^{d-f+1} \left( \frac{k_{ph}}{M_p} \right)^{d-f+1} \quad (d - f \text{ odd}), \quad (3.15)$$

since each factor of  $k/M_p$  must be accompanied by a factor of  $a$  to render the spatial momentum physical. On the other hand, if  $d - f$  is even,  $l_{max}$  as defined above is odd and the term with the highest number of powers of  $k$  is purely imaginary. As we have seen in the previous section, such a term disappears when we add the contribution from the complex

$d - f$	Leading correction	$d - f$	Leading correction
0	$H^2/M_p^2$	4	$H^6 k_{ph}^4/M_p^{10}$
1	$H^2 k_{ph}^2/M_p^4$	5	$H^6 k_{ph}^6/M_p^{12}$
2	$H^4 k_{ph}^2/M_p^6$	6	$H^8 k_{ph}^6/M_p^{14}$
3	$H^4 k_{ph}^4/M_p^8$	7	$H^8 k_{ph}^8/M_p^{16}$

TABLE I: Leading corrections to the gravitational wave two-point functions in the short-wavelength limit.

conjugate graph. Therefore, the leading correction corresponds to the largest *even* value of  $l$ , which turns out to be  $l_{max} = d - f$ , and is therefore given by

$$\delta \langle v^*(\mathbf{k})v(\mathbf{k}) \rangle \propto \frac{1}{2k} \left( \frac{H}{M_p} \right)^{d-f+2} \left( \frac{k_{ph}}{M_p} \right)^{d-f} \quad (d - f \text{ even}) . \quad (3.16)$$

Equations (3.15) and (3.16) represent the main results of this section: they express the leading corrections to the two-point function (in the sub-horizon limit) associated with a generic operator of dimension  $2d + 4$  containing  $2f$  powers of the inflaton field. Since we have assumed an approximate shift invariance of the inflaton, the total number of derivatives,  $2d - 2f + 2$ , must be greater or equal than the number of fields  $2f$ , which in turn implies that  $d \geq f$ . Thus, we can label all the possible corrections with the non-negative index  $d - f$ . Their magnitude is given in Table I for the first eight values of  $d - f$ . Note that corrections with  $d - f = 0$  arise from the operators identified by Weinberg in [4]. The leading momentum-dependent corrections are given by operators with  $d - f = 1$ .

So far, we have calculated the largest possible corrections to the two-point function in the sub-horizon limit given a certain value of  $d - f$ . However, the reader might still wonder whether such terms can be actually obtained from a covariant action. Employing the same technique we used to study the impact of the lowest order terms, it is indeed possible to show—after some rather lengthy calculations—that the following family of covariant terms generates this kind of contributions,

$$\begin{aligned}
 d - f = 0 : & \quad R^{\mu\nu} R_{\mu\nu} \\
 d - f = 1 : & \quad (\nabla^\alpha R^{\mu\nu}) \nabla_\alpha R_{\mu\nu} \\
 d - f = 2 : & \quad (\nabla^\alpha \nabla^\beta R^{\mu\nu}) \nabla_\alpha \nabla_\beta R_{\mu\nu} \\
 & \vdots \quad \vdots \quad .
 \end{aligned} \quad (3.17)$$

It can be also verified that the  $d - f = 1$  term yields a correction to the two-point function proportional to the slow-roll parameter  $\epsilon$ , and given the structure of this family of operators, we anticipate the remaining terms to share the same slow-roll suppression.

### C. The Breakdown of Perturbation Theory

The corrections to the two-point function are functions of two dimensionful parameters,  $H$  and  $k_{ph}$ . For our purposes, it is convenient to organize these corrections in powers of  $k_{ph}$ . Thus, following Table I, and reintroducing the subleading terms that we previously neglected, we find that the two-point function is

$$\langle v^*(\mathbf{k})v(\mathbf{k}) \rangle \approx \frac{1}{2k} \left[ \left( 1 + \alpha_{20} \frac{H^2}{M_p^2} + \dots \right) + \left( \alpha_{22} \frac{H^2}{M_p^2} + \alpha_{42} \frac{H^4}{M_p^4} + \dots \right) \frac{k_{ph}^2}{M_p^2} + \left( \alpha_{44} \frac{H^4}{M_p^4} + \alpha_{64} \frac{H^6}{M_p^6} + \dots \right) \frac{k_{ph}^4}{M_p^4} + \dots \right]. \quad (3.18)$$

The coefficient  $\alpha_{20}$  is of order one, while all the  $\alpha_{nn}$  with  $n \geq 2$  are of order  $\epsilon$ , as the family of covariant terms (3.18) suggests. At the end of Section IV we provide further evidence supporting this claim.

In order for Equation (3.18) to be a valid perturbative expansion, every correction term must be much smaller than one. Because  $\alpha_{20}$  is of order one, this implies the condition

$$\frac{H}{M_p} \ll 1, \quad (3.19)$$

which must hold for all values of  $k_{ph}$ . Equation (3.18) then shows that if condition (3.19) is satisfied, the corrections to the two-point function remain small even for  $k_{ph} \approx M_p$ . In fact, to leading order in  $H/M_p$  we can rewrite equation (3.18) as

$$\langle v^*(\mathbf{k})v(\mathbf{k}) \rangle \approx \frac{1}{2k} \left[ 1 + \alpha_{22} \frac{k_{ph}^2}{\Lambda^2} + \alpha_{44} \frac{k_{ph}^4}{\Lambda^4} + \dots \right], \quad (3.20)$$

where we have introduced the effective cut-off

$$\Lambda \approx \frac{M_p^2}{H}. \quad (3.21)$$

Equations (3.20) and (3.21) are the main result of this article. For  $k_{ph} \ll \Lambda$ , all the corrections are strongly suppressed and can thus be neglected. However, at  $k_{ph} \approx \Lambda$ , all the

corrections become of order  $\epsilon$ , the asymptotic series breaks down, and the effective theory ceases to be valid.

To conclude this section, let us briefly comment on the effects of terms that break the shift symmetry. Because the only difference is that these terms contain undifferentiated scalars, any such correction can be cast as a generally covariant term that respects the symmetry, multiplied by a power of the dimensionless ratio  $\varphi/M_p$ . Hence, these terms introduce corrections to the two-point function of the form we have already discussed, but with coefficients  $\alpha_{ij}$  that can now depend on arbitrary powers of the background field  $\varphi_0$ ,

$$\alpha_{ij} = \alpha_{ij}^{(0)} + \alpha_{ij}^{(1)} \frac{\varphi_0}{M_p} + \alpha_{ij}^{(2)} \left( \frac{\varphi_0}{M_p} \right)^2 + \dots \quad (3.22)$$

Therefore, in the absence of any mechanism or symmetry that keeps the coefficients  $\alpha_{ij}^{(1)}, \alpha_{ij}^{(2)}, \dots$  small (e.g. an approximate shift symmetry), such an expansion loses its validity for  $\varphi_0 > M_p$ , regardless of the value of  $k_{ph}$ . If, on the other hand, equation (3.22) is a sensible expansion, and  $\alpha_{ij}^{(0)}$  is much greater than  $\alpha_{ij}^{(1)}, \alpha_{ij}^{(2)}, \dots$ , then we can effectively assume that the shift-symmetry is exact, and perturbations theory breaks down again at  $k_{ph} \approx \Lambda$ .

#### D. Loop Diagrams

Our analysis so far has concentrated on tree-level diagrams, though loop corrections could also invalidate the short-wavelength expansion. The contribution of loops is obscured by the appearance of divergent momentum integrals. At any order in the derivative expansion it is still possible to cancel these divergences by renormalizing a finite number of parameters, provided that all terms consistent with the symmetries of the theory are included in the Lagrangian [36]. In practice, this cancellation is due to the presence of appropriate counter-terms in the Lagrangian. For this reason, divergent integrals in loop diagrams are rather harmless. They yield corrections of the same structure as tree-level diagrams, modulo a (mild) logarithmic running of their values with scale [37]. Hence, we do not expect this type of contributions to drastically change our conclusions, though we should emphasize that this is just an expectation.

## IV. THE LIMITS OF PERTURBATION THEORY: SCALARS

We now turn our attention to corrections to the two-point function of scalar perturbations. Despite some complications that are particular to this sector, the method developed in the previous section can be easily extended to scalars.

To this end, let us consider the action  $S = S_0 + \lambda S_1$ , where  $S_0$  is the action (2.1) describing a scalar field minimally coupled to Einstein gravity, while  $S_1$  is a generic generally covariant correction suppressed by a coupling  $\lambda \sim 1/M_p^{2d}$ . As we pointed out in the previous section,  $S_1$  generically involves contractions of the Riemann tensor  $R_{\mu\nu\lambda\rho}$  and the covariant derivative  $\nabla_\mu$  as well as the scalar field  $\varphi$ . In order to compute the resulting first order contribution to the two-point function for  $v = a\delta\varphi$ , we insert the perturbed metric (2.9) and the perturbed inflaton field into the action  $S$  and expand up to second order in  $\delta\varphi, \phi$  and  $B$ . We then express the metric perturbations in terms of  $\delta\varphi$  using equations (2.10) and, finally, in terms of  $v$  using the definition (2.11). Note that, even though the relations (2.10) were derived by solving the constraints associated with the unperturbed action for the perturbations  $\delta_2 S_0$ , corrections to these relations due to  $\delta_2 S_1$  do not contribute at first order in  $\lambda$ . To show this, let us expand  $\phi$  and  $B$  in powers of the coupling

$$\phi = \phi_0 + \lambda\phi_1 + \mathcal{O}(\lambda^2), \quad B = B_0 + \lambda B_1 + \mathcal{O}(\lambda^2), \quad (4.1)$$

where  $\phi_0$  and  $B_0$  satisfy the unperturbed relations (2.10a) and (2.10b) respectively. By expanding the full quadratic action for the perturbations  $\delta_2 S[\delta\varphi, \phi, B]$  to first order in  $\lambda$  we obtain

$$\begin{aligned} \delta_2 S[\delta\varphi, \phi, B] &\approx \delta_2 S_0[\delta\varphi, \phi_0, B_0] + \lambda \int \left[ \frac{\delta(\delta_2 S_0)}{\delta\phi} \right]_{\lambda=0} \phi_1 + \lambda \int \left[ \frac{\delta(\delta_2 S_0)}{\delta B} \right]_{\lambda=0} B_1 + \\ &+ \lambda \delta_2 S_1[\delta\varphi, \phi_0, B_0] + \mathcal{O}(\lambda^2). \end{aligned} \quad (4.2)$$

However, the second and the third term vanish because they contain the unperturbed constraint equations evaluated at  $\phi_0$  and  $B_0$ , which by assumption are solutions to the constraints. Therefore, as long as we are interested in first order results, we can safely use the unperturbed solutions  $\phi_0$  and  $B_0$  given by equations (2.10).

Expanding the leading action  $S_0$  to quadratic order in the perturbations, we obtain the free action (2.7). The additional quadratic terms stemming from  $S_1$  must be appropriately contracted expressions containing partial derivatives of the perturbation variable  $v$ , the scale

factor  $a$  and the background field  $\varphi_0$ . However, in the case of scalar perturbations, the field  $v$  can arise from fluctuations of the scalar field,  $\delta\varphi = v/a$ , or from fluctuations of the metric,

$$\delta g^{\mu\nu} = -\frac{a\sqrt{2\epsilon}}{M_p} \left[ 2v_k \delta^\mu{}_0 \delta^\nu{}_0 + \frac{ik_j}{k^2} (v'_k - aHv_k) (\delta^\mu{}_j \delta^\nu{}_0 + \delta^\mu{}_0 \delta^\nu{}_j) \right] \equiv \frac{a\sqrt{2\epsilon} \mathcal{V}^{\mu\nu}}{M_p}. \quad (4.3)$$

Thus, unlike the case of tensor perturbations,  $\delta g^{\mu\nu}$  provides an additional tensor structure that can be used to contract derivatives. We now show that such contractions yield terms where the derivatives acting on  $v$  or  $a$  are contracted with  $\eta^{\mu\nu}$ . This means that the argument in the previous section can be applied to scalar perturbations as well, yielding essentially the same results. We note that terms which contain only fluctuations coming directly from the scalar field do not present this problem, and can be easily written as in equation (3.8).

Let us first consider terms with only one factor of  $\mathcal{V}^{\mu\nu}$ . In this case,  $\mathcal{V}^{\mu\nu}$  can be contracted either with  $\eta_{\mu\nu}$ , leading to  $\mathcal{V}^{\mu\nu} \eta_{\mu\nu} = 2v$ , or with two derivatives  $\partial_\mu \partial_\nu$ , resulting in

$$\partial_\mu \partial_\nu \mathcal{V}^{\mu\nu} = 2 \left( \frac{\partial_\mu \partial^\mu a}{a} v - \frac{\partial_\mu a \partial^\mu a}{a^2} v + \frac{\partial_\mu a \partial^\mu v}{a} \right), \quad (4.4a)$$

$$\partial_\mu a \partial_\nu \mathcal{V}^{\mu\nu} = \partial_\mu a \partial^\mu v + \frac{\partial_\mu a \partial^\mu a}{a} v, \quad (4.4b)$$

$$\partial_\mu \partial_\nu [a, \varphi_0] \mathcal{V}^{\mu\nu} = 2(\partial_\mu \partial^\mu [a, \varphi_0]) v, \quad (4.4c)$$

where, again, the square brackets in the last line mean that the derivatives can act on *one or more* factors of  $a$  or  $\varphi_0$ . Thus, terms with only one factor of  $\mathcal{V}_{\mu\nu}$  do not present any problem since, as anticipated, all the derivatives are contracted with the inverse of the Minkowski metric.

Corrections which contain two factors of  $\mathcal{V}^{\mu\nu}$ , and are not products of terms in (4.4b), can always be recast as

$$\mathcal{V}_{\mu\nu} \mathcal{V}^{\mu\nu} = 2v^2 + \frac{2}{k^2} \left[ \partial_\mu v \partial^\mu v - \frac{\partial_\mu a}{a} \partial^\mu (v^2) + \frac{\partial_\mu a \partial^\mu a}{a^2} v^2 \right], \quad (4.5a)$$

$$\begin{aligned} \partial^\mu \mathcal{V}_{\mu\nu} \partial_\lambda \mathcal{V}^{\lambda\nu} &= -\frac{1}{k^2} \partial_\mu v' \partial^\mu v' + \frac{1}{k^2} \left[ 2 \frac{\partial_\mu a \partial^\mu v}{a} \partial_\nu \partial^\nu v - 2 \left( \frac{\partial_\mu \partial_\nu a}{a} - \frac{\partial_\mu a \partial_\nu a}{a^2} \right) v \partial^\mu \partial^\nu v - \right. \\ &\quad \left. \frac{\partial_\mu a \partial^\mu a}{a^2} \partial_\nu v \partial^\nu v + \left( \frac{\partial_\mu \partial^\mu a}{a} - \frac{\partial_\mu a \partial^\mu a}{a^2} \right) \frac{\partial_\nu a}{a} \partial^\nu (v^2) \right]. \end{aligned} \quad (4.5b)$$

All the terms inside the square brackets become negligible in the sub-horizon limit, since their contribution is suppressed by an extra factor of  $1/k^2$ . The only term in which some derivatives are not contracted with the Minkowski inverse metric is the first one in equation (4.5b). However, the two derivatives with respect to conformal time result in a factor of  $k^2$

which is precisely canceled by the extra factor  $1/k^2$ , and for all practical purposes such a term is equivalent to  $\partial_\mu v \partial^\mu v$ .

Therefore, we have demonstrated that terms quadratic in the scalar fluctuations can be schematically written as in equation (3.8). The remainder of the analysis then proceeds as for tensor perturbations, and effective corrections to scalar perturbations are thus also subdominant in the regime

$$H \ll M_p \quad \text{and} \quad k_{ph} \ll \Lambda \sim \frac{M_p^2}{H}. \quad (4.6)$$

Before concluding, we would like to address again whether the operators that we have considered can be actually obtained from generally covariant terms. In the case of scalar perturbations, it is indeed possible to show—after further rather lengthy calculations—that the following family of covariant terms generates the kind of corrections shown in Table I,

$$\begin{aligned} d - f = 0 : & R^{\mu\nu} (\nabla_\mu \varphi) \nabla_\nu \varphi \\ d - f = 1 : & R^{\mu\nu} (\nabla_\mu \varphi) \nabla_\nu \nabla_\gamma \nabla^\gamma \varphi \\ d - f = 2 : & (\nabla^\alpha \nabla^\beta R^{\mu\nu}) (\nabla_\alpha \nabla_\mu \varphi) \nabla_\beta \nabla_\nu \varphi \\ d - f = 3 : & (\nabla^\alpha \nabla^\beta R^{\mu\nu}) (\nabla_\alpha \nabla_\mu \varphi) \nabla_\beta \nabla_\nu \nabla_\gamma \nabla^\gamma \varphi \\ & \vdots \qquad \vdots \qquad . \end{aligned} \quad (4.7)$$

In order to illustrate how this happens, let us consider for example the  $d - f = 1$  term. It contains, among many other terms a factor

$$a^2 R^{\mu\nu} \partial_\mu \delta\varphi \partial_\nu \partial_\gamma \partial^\gamma \delta\varphi \supset \frac{2\epsilon}{a^6} \partial^\mu a \partial^\nu a \partial_\mu v \partial_\nu \partial_\gamma \partial^\gamma v \sim -\frac{2\epsilon}{a^6} \partial^\mu a \partial^\nu a \partial_\mu \partial_\nu v \square v + \dots \quad (4.8)$$

where, in the last step, we have neglected a subdominant contribution in the short-wavelength limit. The last term in (4.8) indeed generates a correction proportional to  $H^2 k_{ph}^2 / M_p^4$  and it is suppressed by one factor of the slow-roll parameter. It is relatively easy to verify that the corrections generated by the other members of the family (4.7) have the same slow-roll suppression, which strongly supports the assumption we made in the context of tensor perturbations.

## V. CONCLUSIONS

The connection, through cosmological inflation, between physics on the smallest scales, described by quantum field theory, and that on the largest scales in the universe is one of

the most profound aspects of modern cosmology. However, since inflation takes place at such early epochs, and magnifies fluctuations of such small wavelengths, it is important to establish the regime of validity of the usual formalism—that of semiclassical gravity, with quantum field theory assumed valid, and coupled to the minimal Einstein-Hilbert action—at those scales.

On general grounds, we expect the canonical approach to break down at ultra-short distances, where the operators that arise in an effective field theory treatment of the coupled metric-inflaton system become relevant. In this article we have calculated the impact of these higher-dimensional operators on the power spectrum at short wavelengths. In this way, we have been able to determine the regime in which the properties of the perturbations deviate from what is conventionally assumed. From a purely theoretical standpoint, these considerations are important if we are to understand the limits of applicability of cosmological perturbation theory. From an observational standpoint, cosmic microwave background measurements are becoming so precise that we may hope to use them to identify the signatures of new gravitational or field theoretic physics.

Our analysis has focused on tree-level corrections to the spectrum. Because we have essentially considered all possible generally covariant terms in the effective action, we expect to have unveiled the form of all possible corrections that are compatible with the underlying symmetries of the theory. Loop diagrams may yield additional scale-dependent logarithmic corrections neglected in our analysis. In any case, our results indicate that cosmological perturbation theory does not apply all the way to infinitesimally small distances,  $k_{ph} \rightarrow \infty$ , and that, indeed, there is a physical spatial momentum  $\Lambda$  (or a physical length  $1/\Lambda$ ) beyond which cosmological perturbation theory breaks down. The scale at which perturbation theory breaks down at tree-level is

$$\Lambda \sim \frac{M_p^2}{H}, \quad (5.1)$$

which, because of existing limits on the scalar to tensor power spectrum ratio [38], is at least  $10^4$  times the Planck scale.

These results have significant implications for the impact of an effective cut-off  $\Lambda$  on the primordial spectrum of primordial perturbations, which typically is at most of order  $H/\Lambda$  [19]. Substituting the value of  $\Lambda$  we have found, we obtain corrections of the order of  $H^2/M_p^2$ , which are likely to remain unobservable [23]. This value of  $\Lambda$  also solves a problem that was

noticed in [39], namely, that in the presence of a Planckian cut-off, cosmological perturbations do not tend to decay into the Bunch-Davis vacuum (or similar states). In particular, to lowest order in perturbation theory, the transition probability from an excited state into the Bunch-Davis vacuum is significantly less than one for  $\Lambda = M_p$ , but proportionally larger if  $\Lambda$  is given by (5.1). Ultimately, a large decay probability is what justifies the choice of the Bunch-Davies vacuum as the preferred initial state for the perturbations at scales below the cut-off, since, as we have found, our theories lose their validity at momentum scales above the spatial momentum  $\Lambda$ .

### Acknowledgments

The authors thank Robert Brandenberger, Jaume Garriga and Richard Holman for useful comments and conversations. The work of CAP and RP was supported in part by the National Science Foundation (NSF) under grant PHY-0604760. The work of MF and MT was supported in part by the NSF under grants PHY-0354990 and PHY-0653563, by funds from Syracuse University and by Research Corporation.

### APPENDIX A

In this appendix, we show how to integrate by parts every term of the form

$$\partial^{2n+m+p} [a, \varphi_0] \partial^{2q+m+r} v \partial^{2s+p+r} v \quad (\text{A1})$$

in order to express it as a linear combination of terms like

$$\partial^{2j+l} [a, \varphi_0] \partial^{2m+l} v \square v \quad (\text{A2})$$

plus, possibly, a term with no derivatives acting on  $v$ . Notice that, for notational convenience, we have defined  $\square \equiv \partial_\mu \partial^\mu$ . Of course, if the index  $q$  (or  $s$ ) in equation (A1) is not zero, we can easily integrate by parts  $2q + m + r - 2$  ( $2s + p + r - 2$ ) times to get only terms of the form of that in equation (A2). Therefore, in what follows we only consider terms with  $q = s = 0$ . In this case, we can always integrate by parts an appropriate number of times to get only terms for which  $m = p$ . Thus, without loss of generality, we can restrict ourselves to considering terms of the form

$$\partial^{2n+m+p} [a, \varphi_0] \partial^{m+r} v \partial^{p+r} v, \quad (m = p) . \quad (\text{A3})$$

The derivatives acting on  $v$  that are contracted with derivatives acting on  $a$  or  $\varphi_0$  can be systematically eliminated by repeated integrations by parts:

$$\begin{aligned} \partial^{2n+m+p} [a, \varphi_0] \partial^{m+r} v \partial^{p+r} v &\sim -\partial^{2n+m+(p-1)} [a, \varphi_0] \partial^{m+r} v \partial^{(p-1)+r} \square v \\ &+ \frac{1}{2} \partial^{2(n+1)+(m-1)+(p-1)} [a, \varphi_0] \partial^{(m-1)+(r+1)} v \partial^{(p-1)+(r+1)} v , \end{aligned} \quad (\text{A4})$$

where we have denoted equivalence up to integration by parts with  $\sim$ . The first term on the right hand side can be cast in the form (A2) by integrating by parts  $(p-1)+r$  times, while the second one is of the form (A3) with  $n$  and  $r$  ( $p$  and  $m$ ) increased (decreased) by one. By iterating this procedure, we eventually obtain terms of the form

$$\partial^{2n} [a, \varphi_0] \partial^r v \partial^r v , \quad (\text{A5})$$

where now  $n$  and  $r$  have changed. Again, we can integrate by parts and obtain

$$\partial^{2n} [a, \varphi_0] \partial^r v \partial^r v \sim -\partial^{2n} [a, \varphi_0] \partial^{r-1} v \partial^{r-1} \square v + \frac{1}{2} \partial^{2(n+1)} [a, \varphi_0] \partial^{(r-1)} v \partial^{(r-1)} v . \quad (\text{A6})$$

The first term on the right hand side can be re-written as (A2) after  $r-1$  integrations by parts, while the second term has the form (A5) with  $n$  ( $r$ ) increased (decreased) by one. Thus, by repeating this procedure we obtain many terms of the form (A2) and we are eventually left with a term without derivatives acting on  $v$ . This completes our proof.

- [1] A. H. Guth, “The Inflationary Universe: A Possible Solution To The Horizon And Flatness Problems,” Phys. Rev. D **23**, 347 (1981). A. D. Linde, “A New Inflationary Universe Scenario: A Possible Solution Of The Horizon, Flatness, Homogeneity, Isotropy And Primordial Monopole Problems,” Phys. Lett. B **108**, 389 (1982). A. Albrecht and P. J. Steinhardt, “Cosmology For Grand Unified Theories With Radiatively Induced Symmetry Breaking,” Phys. Rev. Lett. **48**, 1220 (1982).
- [2] V. F. Mukhanov and G. V. Chibisov, “Quantum Fluctuation And ‘Nonsingular’ Universe. (In Russian),” JETP Lett. **33**, 532 (1981) [Pisma Zh. Eksp. Teor. Fiz. **33**, 549 (1981)]. S. W. Hawking, “The Development Of Irregularities In A Single Bubble Inflationary Universe,” Phys. Lett. B **115**, 295 (1982). A. H. Guth and S. Y. Pi, “Fluctuations In The New Inflationary Universe,” Phys. Rev. Lett. **49**, 1110 (1982). A. A. Starobinsky, “Dynamics Of Phase Transition In The

- New Inflationary Universe Scenario And Generation Of Perturbations," Phys. Lett. B **117**, 175 (1982). J. M. Bardeen, P. J. Steinhardt and M. S. Turner, "Spontaneous Creation Of Almost Scale - Free Density Perturbations In An Inflationary Universe," Phys. Rev. D **28**, 679 (1983).
- [3] R. H. Brandenberger, "Inflationary cosmology: Progress and problems," arXiv:hep-ph/9910410.
  - [4] S. Weinberg, "Effective Field Theory for Inflation," arXiv:0804.4291 [hep-th].
  - [5] J. Martin and R. H. Brandenberger, "The trans-Planckian problem of inflationary cosmology," Phys. Rev. D **63**, 123501 (2001) [arXiv:hep-th/0005209].
  - [6] R. H. Brandenberger and J. Martin, "The robustness of inflation to changes in super-Planck-scale physics," Mod. Phys. Lett. A **16**, 999 (2001) [arXiv:astro-ph/0005432].
  - [7] J. C. Niemeyer, "Inflation with a high frequency cutoff," Phys. Rev. D **63**, 123502 (2001) [arXiv:astro-ph/0005533].
  - [8] A. Kempf, "Mode generating mechanism in inflation with cutoff," Phys. Rev. D **63**, 083514 (2001) [arXiv:astro-ph/0009209].
  - [9] A. A. Starobinsky, "Robustness of the inflationary perturbation spectrum to trans-Planckian physics," Pisma Zh. Eksp. Teor. Fiz. **73**, 415 (2001) [JETP Lett. **73**, 371 (2001)] [arXiv:astro-ph/0104043].
  - [10] R. Easther, B. R. Greene, W. H. Kinney and G. Shiu, "Inflation as a probe of short distance physics," Phys. Rev. D **64**, 103502 (2001) [arXiv:hep-th/0104102].
  - [11] R. Easther, B. R. Greene, W. H. Kinney and G. Shiu, "Imprints of short distance physics on inflationary cosmology," Phys. Rev. D **67**, 063508 (2003) [arXiv:hep-th/0110226].
  - [12] J. C. Niemeyer and R. Parentani, "Trans-Planckian dispersion and scale-invariance of inflationary perturbations," Phys. Rev. D **64**, 101301 (2001) [arXiv:astro-ph/0101451].
  - [13] A. Kempf and J. C. Niemeyer, "Perturbation spectrum in inflation with cutoff," Phys. Rev. D **64**, 103501 (2001) [arXiv:astro-ph/0103225].
  - [14] R. H. Brandenberger and J. Martin, "On signatures of short distance physics in the cosmic microwave background," Int. J. Mod. Phys. A **17**, 3663 (2002) [arXiv:hep-th/0202142].
  - [15] U. H. Danielsson, "A note on inflation and trans-Planckian physics," Phys. Rev. D **66**, 023511 (2002) [arXiv:hep-th/0203198].
  - [16] R. Easther, B. R. Greene, W. H. Kinney and G. Shiu, "A generic estimate of trans-Planckian

- modifications to the primordial power spectrum in inflation,” Phys. Rev. D **66**, 023518 (2002) [arXiv:hep-th/0204129].
- [17] L. Bergstrom and U. H. Danielsson, “Can MAP and Planck map Planck physics?,” JHEP **0212**, 038 (2002) [arXiv:hep-th/0211006].
- [18] N. Kaloper, M. Kleban, A. E. Lawrence and S. Shenker, “Signatures of short distance physics in the cosmic microwave background,” Phys. Rev. D **66**, 123510 (2002) [arXiv:hep-th/0201158].
- [19] C. Armendariz-Picon and E. A. Lim, “Vacuum choices and the predictions of inflation,” JCAP **0312**, 006 (2003) [arXiv:hep-th/0303103].
- [20] J. Martin and R. Brandenberger, “On the dependence of the spectra of fluctuations in inflationary cosmology on trans-Planckian physics,” Phys. Rev. D **68**, 063513 (2003) [arXiv:hep-th/0305161].
- [21] O. Elgaroy and S. Hannestad, “Can Planck-scale physics be seen in the cosmic microwave background?,” Phys. Rev. D **68**, 123513 (2003) [arXiv:astro-ph/0307011].
- [22] J. Martin and C. Ringeval, “Superimposed Oscillations in the WMAP Data?,” Phys. Rev. D **69**, 083515 (2004) [arXiv:astro-ph/0310382].
- [23] T. Okamoto and E. A. Lim, “Constraining Cut-off Physics in the Cosmic Microwave Background,” Phys. Rev. D **69**, 083519 (2004) [arXiv:astro-ph/0312284].
- [24] S. Shankaranarayanan and L. Sriramkumar, “Trans-Planckian corrections to the primordial spectrum in the infra-red and the ultra-violet,” Phys. Rev. D **70**, 123520 (2004) [arXiv:hep-th/0403236].
- [25] J. Martin and C. Ringeval, “Exploring the superimposed oscillations parameter space,” JCAP **0501**, 007 (2005) [arXiv:hep-ph/0405249].
- [26] K. Schalm, G. Shiu and J. P. van der Schaar, “Decoupling in an expanding universe: Boundary RG-flow affects initial conditions for inflation,” JHEP **0404**, 076 (2004) [arXiv:hep-th/0401164].
- [27] B. R. Greene, K. Schalm, G. Shiu and J. P. van der Schaar, “Decoupling in an expanding universe: Backreaction barely constrains short distance effects in the CMB,” JCAP **0502**, 001 (2005) [arXiv:hep-th/0411217].
- [28] K. Schalm, G. Shiu and J. P. van der Schaar, “The cosmological vacuum ambiguity, effective actions, and trans-Planckian effects in inflation,” AIP Conf. Proc. **743**, 362 (2005) [arXiv:hep-th/0412288].

- [29] R. Easther, W. H. Kinney and H. Peiris, “Observing trans-Planckian signatures in the cosmic microwave background,” *JCAP* **0505**, 009 (2005) [arXiv:astro-ph/0412613].
- [30] B. Greene, K. Schalm, J. P. van der Schaar and G. Shiu, “Extracting new physics from the CMB,” *In the Proceedings of 22nd Texas Symposium on Relativistic Astrophysics at Stanford University, Stanford, California, 13-17 Dec 2004, pp 0001* arXiv:astro-ph/0503458.
- [31] R. Holman and A. J. Tolley, “Enhanced Non-Gaussianity from Excited Initial States,” arXiv:0710.1302 [hep-th].
- [32] W. Hu, “Covariant Linear Perturbation Formalism,” arXiv:astro-ph/0402060.
- [33] V. F. Mukhanov, “Quantum Theory of Gauge Invariant Cosmological Perturbations,” Sov. Phys. JETP **67**, 1297 (1988) [Zh. Eksp. Teor. Fiz. **94N7**, 1 (1988)].
- [34] S. Weinberg, “Quantum contributions to cosmological correlations,” Phys. Rev. D **72**, 043514 (2005) [arXiv:hep-th/0506236].
- [35] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan and L. Senatore, “The Effective Field Theory of Inflation,” *JHEP* **0803**, 014 (2008) [arXiv:0709.0293 [hep-th]].
- [36] S. Weinberg, “The Quantum theory of fields. Vol. 1: Foundations”, *Cambridge Univ. Pr.* (1995) 609 p
- [37] J. F. Donoghue, “General Relativity As An Effective Field Theory: The Leading Quantum Corrections,” Phys. Rev. D **50**, 3874 (1994) [arXiv:gr-qc/9405057].
- [38] D. N. Spergel *et al.* [WMAP Collaboration], “Wilkinson Microwave Anisotropy Probe (WMAP) three year results: Implications for cosmology,” *Astrophys. J. Suppl.* **170**, 377 (2007) [arXiv:astro-ph/0603449].
- [39] C. Armendariz-Picon, “Why should primordial perturbations be in a vacuum state?,” *JCAP* **0702**, 031 (2007) [arXiv:astro-ph/0612288].